Memorandum

Date: February 26, 2019

To: Bob Van Dyk, Wild Salmon Center

From: Don L. Stevens, Jr.

RE: Comments ON "Forest practices compliance audit: 2017Annual Report"

Introduction

My name is Don Stevens. I am a retired Senior Research Professor from Oregon State University's Department of Statistics. I have over 30 years of experience in applying quantitative methods to issues arising in the environmental, biological, and physical sciences. For over ten years, I was engaged in developing the statistical sampling theory supporting the USEPA's Environmental Monitoring and Assessment Programs spatially balanced probability sampling, and simultaneously applying that theory to designing samples of a variety of environmental resources, such as lakes and streams, forests, and estuaries. I am a fellow of the American Statistical Association, and was awarded the Distinguished Achievement Medal by the Section on Statistics and the Environment of the ASA. I am a past president of the International Environmetrics Society and an elected member of The International Statistical Institute. After retiring; I was an active consultant on environmental monitoring and sampling design. Former clients include the San Francisco Estuary Institute, the Oregon Department of Fish and Wildlife (ODFW), California Department of Fish and Game, and Australia's Commonwealth Science and Industrial Research Organization's Environmental Informatics group.

You asked me to comment on the design and validity of the conclusions of the 2017 Compliance Audit. Briefly, the audit was intended to be a rigorous statistical assessment of compliance rates with the Oregon Forest Practices Act (FPA). A sample of "forest operation sites" was selected from the Forest Activity Electronic Reporting and Notifications System (FERNS) and was used to assess compliance with FPA rules.

A primary advantage of a properly designed and executed statistical survey is the ability to assess precision, e.g., by calculating confidence intervals. Unfortunately, the selection process used in the survey has severe flaws which were compounded by the analysis. These flaws are sufficiently severe to preclude any defendable assessment of compliance rates. The audit program should not proceed without the advice of a qualified statistician with experience in survey design.

Even if the design had been correctly implemented, the analysis used is not amenable to calculating precision or confidence interval estimates. Compliance rate was calculated for a stratum by dividing total number of non-compliant rule applications by total number of rule applications. There is certainly covariance among rule applications on a site so that a legitimate

confidence is difficult or impossible to calculate. The usual binomial based computation does not apply.

Comments of Study Design

The first steps in selecting a proper statistical sample are to define the population to be sampled and to identify a population frame. A good population frame includes or covers every unit in the population and does not include any units not in the target population. This audit used FERNS as the population frame. FERNS satisfies the first requirement, but not the second: the report notes that of 345 sites initially selected from FERNS, only 155 were deemed suitable. Over-coverage is not in itself a fatal flaw but appropriate methods must be followed to select a legitimate sample, and the methods need to be carried through to the analysis.

The sample was stratified into three areas with sample sizes in each area chosen "proportional to the total acreage for which notifications were received during the sample interval". There is no justification for using total acreage as a stratifying parameter, nor is there any indication that any reporting element is based on acreage. Instead, the reporting is based entirely on proportion, (number of non-compliant sites divided by total number of sites). An additional stratification by ownership class was carried out. It is not stated whether the "stratification" by ownership class was a proper stratification or whether it was simply a sub-sample of the existing overall sample.

Details of the stratification process are not provided, but it does not appear to have been done correctly. The steps that should have been followed are: (1) identify the total number of units in the target population that are in each stratum, and (2) select independent samples from each stratum. This information is necessary to correctly analyze a stratified sample because the stratification introduces variable weights in the sample. Bias results if the variation is ignored in the analysis.

Every site in the initial draw should be cross-classified as in Table 1, i.e., by area and landowner class. Once this is done, then a target sample size should be specified for each stratum and samples for each stratum should be drawn independently.

The most serious problem with the study is the high rate of non-response. Of the 354 sites in the initial sample, 111 were classified as non-target and excluded; sampling permission was refused at 40 sites, and no response was received for 79 sites (for a total of 119 sites for which no response was received). Permission to sample was received for only 115 sites. This provides an enormous potential for non-response bias.

When data are missing completely at random (MCAR), the reason for the lack of response is not associated with the outcome of interest, any related covariates, or the survey design. The data are considered a random subsample of the intended sample and analysis can be conducted using standard approaches on the reduced data set. When data are missing at random (MAR), the missingness mechanism is related to levels of one or more covariates. Within these levels, the

missingness is MCAR. Adjustments made within the levels of these covariates provide unbiased inference, but the variance must be adjusted to account for the additional uncertainty. When data are not missing at random (NMAR), then the outcome of interest is related to the missingness mechanism (e.g. poor populations on private land or landowner access refusal) and information from the sample is insufficient to correct the bias. NMAR generally requires information on at least a subset of missing sites or an assumed model.

In this case, the analysis proceeded under the MCAR assumption that the non-response sites would have the same compliance rate as the sampled sites. That assumption does not seem tenable. A reasonable expectation is that the non-compliance rate in the non-response sites is higher because the landowners do not want the site evaluated. A non-response rate of over 50% means that the results of the survey are not credible. There are some techniques that might have helped to reduce the non-response rate. For example, extreme efforts to obtain a response from a random sample of the non-responders could have provided some insight into the homogeneity of responders and non-responders. Extreme efforts could entail multiple contact attempts, in-person visits, or offering rewards for responding.

The report notes that "When a site was found to be unsuitable, or landowner permission could not be obtained, replacement sites were chosen using the random process described previously." It is possible to design a rigorous random sample that allows for site replacement, but this cannot be done by simply selecting more samples. The process must begin with the initial selection by "over-sampling", that is, selecting more samples that required for the final sample.

Comments on analysis

Most importantly, the analysis of a stratified random sample MUST take into account the stratification. Stratification introduces differential weight between strata, and this leads to bias if not accommodated in the analysis. Normally, results are calculated within strata, and then combined to obtain population level results. In particular, it is wrong to simply pool all observations to calculate an overall compliance rate. Observations of multiple applications within a site are unlikely to be independent.

A more basic concern is the definition of compliance rate itself. Compliance rates were calculated based on the total number of potential rule applications for a given stratification (e.g., by Area, ownership class). The number of non-compliant applications was then divided by this number. Aside from the need to do a stratified analysis, there are serious problems with this approach. A memo from Brenda McComb to Lena Tucker mentioned pseudo replication and confidence interval calculation. These are both legitimate concerns. Almost certainly, the rule applications on a site are not independent, and cannot be analyzed as if they were. In particular, the ratio of non-compliant to total rule applications cannot be treated as binomial to calculate

confidence. In fact, obtaining a valid confidence interval for a rate calculated as in the report would be extremely complicated or impossible, even if the audit had been properly designed.

Some serious consideration needs to be devoted to developing an appropriate metric for a composite compliance rate that can be applied in a stratified design. A possible approach would be to derive an aggregate measure of compliance at the site level, and then combine over sites to obtain a population-level metric. This approach was used in ODF 2002. With a site-level metric of compliance, then confidence intervals can be calculated using the standard stratified sample formulae.

Reference

ODF 2002. Oregon Department of Forestry, 2002. Best Management Practices Compliance Monitoring Project: Final Report. April 2002. 75 pp.
http://www.oregon.gov/ODF/Documents/WorkingForests/BMPComplianceReport.pdf